

# Local Partial Clique Covers for Index Coding

Abhishek Agarwal and Arya Mazumdar

## Abstract

Index coding, or broadcasting with side information, is a network coding problem of most fundamental importance. In this problem, given a directed graph, each vertex represents a user with a need of information, and the neighborhood of each vertex represents the side information availability to that user. The aim is to find an encoding to minimum number of bits (optimal rate) that, when broadcasted, will be sufficient to the need of every user. Not only the optimal rate is intractable, but it is also very hard to characterize with some other well-studied graph parameter or with a simpler formulation, such as a linear program. Recently there have been a series of works that address this question and provide explicit schemes for index coding as the optimal value of a linear program with rate given by well-studied properties such as local chromatic number or partial clique-covering number. There has been a recent attempt to combine these existing notions of local chromatic number and partial clique covering into a unified notion denoted as the *local partial clique cover* (Arbabjolfaei and Kim, 2014).

We present a generalized novel upper-bound (encoding scheme) - in the form of the minimum value of a linear program - for optimal index coding. Our bound also combines the notions of local chromatic number and partial clique covering into a new definition of the local partial clique cover, which outperforms both the previous bounds, as well as beats the previous attempt to combination.

Further, we look at the upper bound derived recently by Thapa et al., 2015, and extend their  $n$ -GIC (Generalized Interlinked Cycle) construction to  $(k, n)$ -GIC graphs, which are a generalization of  $k$ -partial cliques.

## I. INTRODUCTION

Index coding is a multiuser communication problem in which the broadcaster reduces the total number of transmissions by taking advantage of the fact that a user may have side information about other user's data. We consider the index coding problem formulation of [3]. Suppose, a set of users are assigned bijectively to a set of vectors that they want to know. However, instead of knowing the assigned vector, each knows a subset of other vectors. This scenario can be depicted by a so-called side-information graph where each vertex represents a user and there is an edge from user A to user B, if A knows the vector assigned to B. Given this graph, how much information should a broadcaster transmit, such that each vertex can deduce its assigned vector?

Consider a group of  $n$  users with the side information represented by the directed graph  $\mathcal{G}$ , such that the out-neighbors of vertex  $v_i \in V(\mathcal{G}) = \{v_1, v_2, \dots, v_n\}$  denote the nodes whose data is available to  $v_i$ . The data requested by node  $v_i$  is represented by a vector  $\mathbf{x}_i \in \mathbb{F}_q^\ell$ . Let  $N(v, \mathcal{G}) \subseteq V(\mathcal{G}) \setminus \{v\}$  denote the out-neighbors of vertex  $v$  ( $v \in \overline{N(v, \mathcal{G})}$ ). For example, an index coding scenario has been depicted in fig. 1. The side information graph for this scenario is given in fig. 2. In the graph  $\mathcal{G}$  of

Abhishek Agarwal is with the Department of Electrical and Computer Engineering, University of Minnesota, Minneapolis, MN 55455. Arya Mazumdar is with the Computer Science Department of University of Massachusetts, Amherst, MA 01003, and was with the University of Minnesota. email: abhiag@umn.edu, arya@cs.umass.edu. Research supported by NSF grants CCF 1318093 and CAREER CCF 1453121.

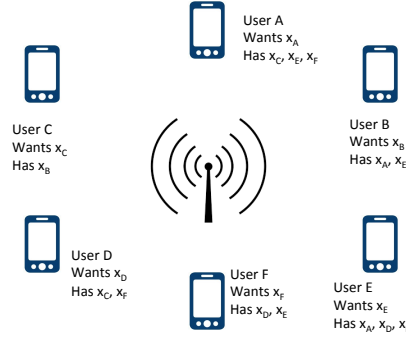


Fig. 1: Example of an Index Coding problem. The corresponding side-information graph is depicted in fig. 2

fig. 2,  $N(A, \mathcal{G}) = \{C, E, F\}$ ,  $N(B, \mathcal{G}) = \{A, E\}$ ,  $N(C, \mathcal{G}) = \{B\}$ ,  $N(D, \mathcal{G}) = \{C, F\}$ ,  $N(E, \mathcal{G}) = \{A, D, F\}$ ,  $N(F, \mathcal{G}) = \{D, E\}$ . The aim is to broadcast the minimum amount of data (in-terms of  $\mathbb{F}_q^\ell$ -symbols) such that vertex  $v_i$  is able to reconstruct  $\mathbf{x}_i$  from  $\mathbf{x}_{N(v_i, \mathcal{G})}$  (vectors assigned to the neighbors) and the broadcast for all  $i \in \{1, \dots, n\}$ . The amount of broadcasted data is referred to as the *broadcast rate* of the index coding problem.

This simple formulation attracted substantial attention recently [1], [5], [12]. In particular, it has been shown that any network coding problem can be cast as an index coding problem [6], [7]. Furthermore, index coding has been identified as the hardest instance of all of network coding [11]. For up to  $n = 5$ , the optimal broadcast rates of all index coding problems have been tabulated in [10].

If the recovery functions for the nodes, as well as the encoding function at the broadcaster are all linear over  $\mathbb{F}_q$ , then the optimal (minimum) broadcasting rate is given by a quantity called the *minrank* of the graph [3], [9]. In general, minrank is hard to compute for arbitrary graphs, and also does not give an achievable scheme in terms of simpler well known graph parameters.

There have been a series of works that aim to characterize the index coding broadcast rate in terms of other possibly intractable graph parameters or cast the index coding rate as the optimal solution to a linear program. To start with, an immediate achievable scheme in terms of the *clique cover* number of the graph or the *chromatic number* of the complementary graph was provided in [4]. Their simple idea is to partition the graph into minimum number of vertex-disjoint cliques. And then for each clique, the broadcaster transmits only once, that is the sum (as  $\mathbb{F}_q$ -vectors) of the data in all the vertices of the clique. This provides an index coding solution equal to the chromatic number of the complementary graph.

The above approach then was extended in [14]. In particular, in [14] an interference alignment perspective of the linear index coding problem was given. It was shown that, it is possible to achieve an index coding solution with number of transmissions equal to the *local chromatic number* [8] of the graph. There also exists an orthogonal approach to generalize the clique covering bound (or the chromatic number bound). In this approach, appearing as early as in [4], one finds a *partial clique cover* (covering with subgraphs instead of cliques) and uses maximum-distance separable (MDS) codes for each of the partial cliques.

More recently, in [2], the above two approaches were merged and a combined linear program solution was proposed that we discuss later. In this paper, we show that these two orthogonal approaches can be merged more effectively, and provide a modified definition of *local partial clique cover*. This lead to better achievability bound for linear index codes than any previous approaches.

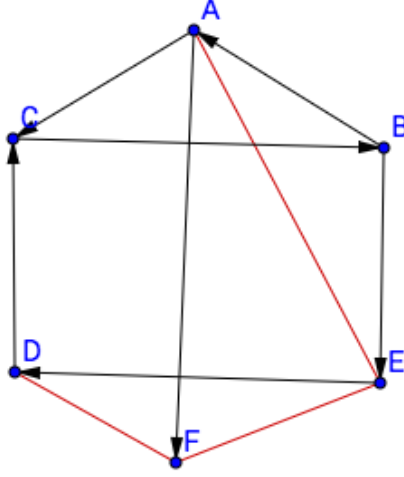


Fig. 2: Example of an Index Coding problem where the fractional local chromatic number and the fractional partial clique number based index codes are strictly larger than the proposed scheme for the fractional case. The undirected edges denote bidirectional edges.

In addition, recently another approach to provide achievable index coding schemes is shown in [15], where the graph is covered by structures called *generalized interlinked cycles (GIC)*. We are also able to generalize this method in the way of partial clique covers.

The paper is organized as follows. In section II we define the necessary graph parameters, describe the previous schemes and narrate our main theorems. In this section we also give some concrete examples to show the better performance of our solution. In section III we provide some further results by showing a generalization of the GIC scheme of [15]. In section IV we describe our main achievability scheme (and hence the proof of the main result) that outperforms the previous results. This is followed up in section V where we provide a linear programming (LP) relaxation. Proofs and the example have been shortened for space constraints.

## II. PRELIMINARIES AND MAIN RESULT

### A. Notations

The following set of notations will be useful throughout the paper, and we provide it at the outset for easy reference.

- $[n] \equiv \{1, 2, \dots, n\}$
- $[m, n] \equiv \{m, m+1, \dots, n\}, m \leq n$
- Complement of set  $A$ , is denoted by  $\bar{A}$
- For a graph  $\mathcal{G}$ ,  $V(\mathcal{G})$  and  $E(\mathcal{G})$  denote the vertex and the edge set of the graph, respectively.  $\bar{\mathcal{G}}$  denotes the directed complement of  $\mathcal{G}$
- For any set  $A = \{i_1, i_2, \dots, i_r\} \subseteq [n]$  and set of vectors  $\{\mathbf{v}_i\}_{i \in [n]}$ ,  $\mathbf{v}_A$  denotes the set  $\{\mathbf{v}_j\}_{j \in A}$  and  $\mathbf{v}_{[A]}$  denotes the matrix  $[\mathbf{v}_{i_1} \ \mathbf{v}_{i_2} \ \dots \ \mathbf{v}_{i_r}]$ . For a matrix  $G \in \mathbb{F}^{k \times n}$ ,  $G_{[B]}$  denotes the sub-matrix of  $G$  constructed from the columns of  $G$  corresponding to  $B \subseteq [n]$ .
- For a graph  $\mathcal{G}(V, E)$  and a set  $S \subseteq V$ ,  $\mathcal{G}|_S$  denotes the subgraph induced by  $S$ .

- For a graph  $\mathcal{G}(V, E)$ ,  $N(v, \mathcal{G}) \subseteq V(\mathcal{G}) \setminus \{v\}$  denotes the set of out-neighbors of  $v \in V$ .
- $I(\text{expr})$  denotes the indicator function for  $\text{expr}$ .
- An  $m \times n$  matrix,  $m \leq n$  is MDS if any  $m$  columns of the matrix are linearly independent.
- A linear code is called MDS code, if the generator (or the parity-check) matrix of the code is MDS. An  $[n, k]$ -MDS code implies the code is of length  $n$  and dimension  $k$ .
- For a matrix  $M$ ,  $M(i, j)$  denotes the entry at  $i^{\text{th}}$  row and  $j^{\text{th}}$  column.

We will need the following definitions.

**Definition 1.** For a directed graph  $\mathcal{G}$ , a coloring of indices is proper if for every vertex  $v$  of  $\mathcal{G}$  the color any of its out-neighbors,  $N(v, \mathcal{G})$ , is different from the color of vertex  $v$ . The minimum number of colors in a proper coloring is the chromatic number  $\chi$ .

**Definition 2.** The local chromatic number  $\chi_l$  of a directed graph  $\mathcal{G}$  is the maximum number of colors in any out-neighborhood (including itself) of a vertex, minimized over all proper colorings of the undirected graph obtained from  $\mathcal{G}$  by ignoring the orientation of edges in  $\mathcal{G}$ .

The following definition of index coding and the optimal broadcast rate will be used throughout the paper. In the index coding problem, a directed side information graph  $\mathcal{G}(V, E)$  is given. Each vertex  $v_i \in V$  represents a receiver that is interested in knowing a uniform random vector  $\mathbf{x}_i \in \mathbb{F}_q^\ell$ . The receiver at  $v$  knows the values of the variables  $\mathbf{x}_u, u \in N(v, \mathcal{G})$ .

**Definition 3** (Index coding). An index code  $\mathcal{C}$  for  $\mathbb{F}_q^{\ell \times n}$  with side information graph  $\mathcal{G}(V, E), V = \{v_1, v_2, \dots, v_n\}$ , is a set of codewords (vectors) in  $\mathbb{F}_q^{\ell \cdot n_0}$  together with:

- 1) An encoding function  $f$  mapping inputs in  $\mathbb{F}_q^{\ell \times n}$  to codewords, and
- 2) A set of deterministic decoding functions  $g_1, \dots, g_n$  such that  $g_i\left(f(\mathbf{x}_1, \dots, \mathbf{x}_n), \{\mathbf{x}_j : j \in N(v_i, \mathcal{G})\}\right) = \mathbf{x}_i$  for every  $i = 1, \dots, n$ .

The encoding and decoding functions depend on  $\mathcal{G}$ . The number  $n_0$  is called the broadcast rate of the index code  $\mathcal{C}$  and  $\ell$  is called the size of a node. Note that  $n_0$  can be a non-integer as long as  $n_0 \cdot \ell$  is integer. The minimum value of  $n_0$  over all possible index codes and all possible  $\ell$  is called the optimal broadcast rate.

### B. Prior work

An *interference alignment* perspective to the graph coloring approach described in the introduction was presented in [14] by assigning a vector  $\mathbf{v}_i$  to  $v_i$  such that,

$$\mathbf{v}_i \notin \text{span}(\mathbf{v}_{\overline{N(v_i, \mathcal{G}) \setminus \{v_i\}}}). \quad (1)$$

From the interference alignment perspective,  $\overline{N(v, \mathcal{G}) \setminus \{v\}}$  are the interfering set of indices for user  $v$ . Let  $G = [\mathbf{v}_1 \mathbf{v}_2 \dots \mathbf{v}_n]$ . Then, the index code (the broadcaster transmission) is given by  $G \cdot [\mathbf{x}_1 \mathbf{x}_2 \dots \mathbf{x}_n]_{\ell \times n}^T \in \mathbb{F}_q^{\ell \times \text{rank}(G)}$ . It can be seen that each node  $v_i$  can recover  $\mathbf{x}_i$  from the index code because of eq. (1). We also sometime denote the vector assigned to node  $v \in V(\mathcal{G})$  as  $\mathbf{v}(v)$ .

It is possible to satisfy the requirements in eq. (1) if  $\text{span}(\mathbf{v}_{[n]}) = n$ . The goal is to minimize the dimension of  $\text{span}(\mathbf{v}_{[n]})$ . One solution to this problem is to find a proper coloring (see, definition 1) of the vertices in  $\overline{\mathcal{G}}$  and assign orthonormal vectors to each color (the same vector is assigned to all vertices with the same color). Thus, an achievable solution is given by the chromatic number of  $\overline{\mathcal{G}}$ ,  $\chi(\overline{\mathcal{G}})$ .

One way to improve the upper-bound  $\chi(\bar{\mathcal{G}})$  above is to use an  $[n, k]$ -MDS code with  $k = \chi_l(\bar{\mathcal{G}})$ . The local chromatic number of  $\bar{\mathcal{G}}$  can be given by the following integer program (IP),

$$\begin{aligned} & \text{minimize} && t \\ & \text{subject to} && \sum_{\mathcal{S} \in \mathcal{K}_0: \mathcal{S} \not\subseteq N(v, \mathcal{G})} \rho_{\mathcal{S}} \leq t, \quad v \in V(\mathcal{G}) \\ & \text{and} && \sum_{\mathcal{S} \in \mathcal{K}_0: v \in \mathcal{S}} \rho_{\mathcal{S}} \geq 1, \quad v \in V(\mathcal{G}) \\ & && \rho_{\mathcal{S}} \in \{0, 1\}, \quad \mathcal{S} \in \mathcal{K}_0 \end{aligned} \quad (2)$$

where  $\mathcal{K}_0$  denotes the set of cliques in the directed graph  $\mathcal{G}$ . Denote  $\rho_{\mathcal{S}}^*$  as the optimal solution to the above IP. Note that there exists a proper coloring of  $\bar{\mathcal{G}}$  corresponding to the IP in eq. (2). We use the generator matrix of a  $[\sum_{\mathcal{S} \in \mathcal{K}_0} \rho_{\mathcal{S}}^*, \chi_l(\bar{\mathcal{G}})]$ -MDS code to assign a column vector to each color in the proper coloring. In [14], it was shown that this scheme works for every directed graph by assigning each vertex a vector corresponding to its color in the proper coloring of eq. (2).

**Theorem 1** (Shanmugam et al. [14]). *The minimum broadcast rate of an index coding problem is upper bounded by the optimal value of the integer program of eq. (2).*

Note that, a coloring of  $\bar{\mathcal{G}}$  is equivalent to a clique covering of  $\mathcal{G}$ . Thus, another approach to finding an index code is to find a *partial clique* (definition 4) cover of  $\mathcal{G}$  [4].

**Definition 4.** A directed graph  $\mathcal{G}$  on  $n$  vertices is a  $k$ -partial clique if every vertex has at least  $n - 1 - k$  out-neighbors and there is at least one vertex in  $\mathcal{G}$  with exactly  $n - k - 1$  out-neighbors.

In each of the  $k_{\mathcal{S}}$ -partial clique  $\mathcal{S}$ , one might use a parity check matrix of a Reed-Solomon code that can correct  $k_{\mathcal{S}} + 1$  erasures. The broadcaster finds the syndrome of the symbols at the vertices. The dimension of the syndrome is  $k_{\mathcal{S}} + 1$ . Thus, the broadcast rate of the index coding problem is given by the following integer program,

$$\begin{aligned} & \text{minimize} && \sum_{\mathcal{S} \in \mathcal{K}} (k_{\mathcal{S}} + 1) \rho_{\mathcal{S}} \\ & \text{subject to} && \sum_{\mathcal{S} \in \mathcal{K}: v \in \mathcal{S}} \rho_{\mathcal{S}} \geq 1, \quad v \in V(\mathcal{G}) \\ & && \rho_{\mathcal{S}} \in \{0, 1\}, \quad \mathcal{S} \in \mathcal{K} \end{aligned} \quad (3)$$

where  $\mathcal{K}$  denotes the power set of vertices in  $\mathcal{G}$  and  $\mathcal{S} \in \mathcal{K}$  is a  $k_{\mathcal{S}}$ -partial clique.

Subsequently in this paper we combine the schemes in [14] and [4] to derive an achievability scheme that would give better results than both of these existing schemes. Recently, in an effort to combination, Arbajjolfaei and Kim [2] proposes the following result.

**Theorem 2** (Arbajjolfaei and Kim [2]). *The minimum broadcast rate of an index coding problem on the side information graph  $\mathcal{G}$  is upper bounded by the optimal value of the following integer program,*

$$\begin{aligned} & \text{minimize} && t \\ & \text{subject to} && \sum_{\mathcal{S} \in \mathcal{K}: \mathcal{S} \not\subseteq N(v, \mathcal{G})} (k_{\mathcal{S}} + 1) \rho_{\mathcal{S}} \leq t, \quad v \in V(\mathcal{G}) \\ & \text{and} && \sum_{\mathcal{S} \in \mathcal{K}: v \in \mathcal{S}} \rho_{\mathcal{S}} \geq 1, \quad v \in V(\mathcal{G}) \\ & && \rho_{\mathcal{S}} \in \{0, 1\}, \quad \mathcal{S} \in \mathcal{K}. \end{aligned} \quad (4)$$

We provide a stronger bound than eq. (4) (see theorem 3 below). Our bound requires a significantly involved achievability scheme.

There is another approach to combine the above two methods of local chromatic number and partial clique covering proposed in [13]. The approach in [13] uses the local chromatic number on partitions of the side information graph. The partitions are decided such that the sum of the local chromatic number of the complement graph of the partitions is minimized. However for the unicast index coding problem that we consider, Theorem 2 can be shown to outperform this scheme of [13].

Very recently another technique for index coding achievability was proposed in [15]. We refrain from separately describing the scheme here. Instead in section III, we provide a generalization of their scheme. The generalization is similar in spirit to that of  $k$ -partial clique covering from clique covering.

### C. Our main results

Our main result is given by the following theorem.

**Theorem 3.** *The minimum broadcast rate of an index coding problem on the side information graph  $\mathcal{G}$  is upper bounded by the optimum value of the following integer program.*

$$\begin{aligned} & \text{minimize } t \\ & \text{subject to } \sum_{\mathcal{S} \in \mathcal{K}} \min\{|\mathcal{S} \cap \overline{N(v, \mathcal{G})}|, (k_{\mathcal{S}} + 1)\} \rho_{\mathcal{S}} \leq t, \end{aligned} \quad (5a)$$

$$v \in V(\mathcal{G})$$

$$\text{and } \sum_{\mathcal{S} \in \mathcal{K}: v \in \mathcal{S}} \rho_{\mathcal{S}} \geq 1, \quad v \in V(\mathcal{G}) \quad (5b)$$

$$\rho_{\mathcal{S}} \in \{0, 1\}, \quad \mathcal{S} \in \mathcal{K}. \quad (5c)$$

Any vertex subgraph of a  $k$ -partial clique is also a  $k$ -partial clique. Thus, the IP in eq. (5) finds a partition of  $\mathcal{G}$  into partial cliques i.e. the inequality in eq. (5b) can be replaced by equality. We call the optimal solution to eq. (5) the *local partial clique cover number*.

To construct a scheme to achieve the bound in theorem 3 we assign vectors to each vertex  $u \in V(\mathcal{G})$ ,  $\mathbf{v}_u$ , such that the condition in eq. (1) is satisfied. And for each partial clique we use the column vectors of a generator matrix of an  $(|\mathcal{S}|, k_{\mathcal{S}} + 1)$ -MDS code. The length of the index code is equal to the span of the vectors  $\{\mathbf{v}_u\}_{u \in V(\mathcal{G})}$ . The detailed scheme is described in section IV.

The term  $\min\{|\mathcal{S} \cap \overline{N(v, \mathcal{G})}|, (k_{\mathcal{S}} + 1)\}$  in eq. (5a) denotes the span of the non-neighbors of vertex  $v$  in the partition  $\mathcal{S}$ . Note that, we count at most  $(k_{\mathcal{S}} + 1)$  non-neighbors in a  $k_{\mathcal{S}}$ -partial clique because the span of the vectors corresponding to  $\mathcal{S}$  is  $k_{\mathcal{S}} + 1$ .

To find the local chromatic number, we partition the graph into cliques and minimize the maximum number of partitions in the neighborhood of any vertex in the complementary side information graph,  $\bar{\mathcal{G}}$ . In theorem 3, we replace the clique cover with a partial clique cover and minimize the maximum span of the vectors in the neighborhood of any vertex in  $\bar{\mathcal{G}}$ . This gives a generalization of the local chromatic number of  $\bar{\mathcal{G}}$  for partial clique cover, and reduces to  $\chi_l(\mathcal{G})$  when only 0-partial cliques are used. Since a clique cover is the special case of a  $k$ -partial cover (a clique is a 0-partial clique), eq. (5) trivially outperforms eqs. (2) and (3).

The change in eq. (5) from eq. (4) is introduction of the term  $|\mathcal{S} \cap \overline{N(v, \mathcal{G})}|$  in the optimization under the summation. Although the set over which the summation is performed seemed to have expanded, it is

not the case. Notice that, if  $S \subseteq N(v, \mathcal{G})$  then  $|\mathcal{S} \cap \overline{N(v, \mathcal{G})}| = 0$ . This also shows that our scheme is at least as good as eq. (4). However the achievability scheme becomes significantly complicated because of the introduction of the term under the summation.

In comparison, for the linear program in eq. (4), a simpler achievability scheme is possible and we describe it in Appendix A for completeness. Note that, this simple achievability scheme was not mentioned in [2].

We can consider the fractional relaxation of the IP in eq. (5) and provide an even better upper-bound.

$$\begin{aligned}
& \text{minimize} && t \\
& \text{subject to} && \sum_{S \in \mathcal{K}} \min\{|\mathcal{S} \cap \overline{N(v, \mathcal{G})}|, (k_S + 1)\} \rho_S \leq t, \\
& && v \in V(\mathcal{G}) \\
& \text{and} && \sum_{S \in \mathcal{K}: v \in S} \rho_S \geq 1, \quad v \in V(\mathcal{G}) \\
& && \rho_S \in [0, 1], \quad S \in \mathcal{K}.
\end{aligned} \tag{6}$$

**Theorem 4** (Linear programming relaxation). *The minimum broadcast rate of an index coding problem on the side information graph  $\mathcal{G}$  is at the most the optimum value of the linear programming of eq. (6).*

The necessary adjustments required for the proof of this theorem is provided in section V.

Following the ideas of [2], we can further tighten the achievability bounds by considering recursive linear program.

**Theorem 5** (Recursive LP). *Let  $IC_{FLP}(\mathcal{G})$  denote the optimized value of the linear program below for graph  $\mathcal{G}$ :*

$$\begin{aligned}
& \text{minimize} && t \\
& \text{subject to} && \sum_{S \in \mathcal{K}} \min\{|\mathcal{S} \cap \overline{N(v, \mathcal{G})}|, IC_{FLP}(\mathcal{G}|_S)\} \rho_S \leq t, \\
& && v \in V(\mathcal{G}) \\
& \text{and} && \sum_{S \in \mathcal{K}: v \in S} \rho_S \geq 1, \quad v \in V(\mathcal{G}) \\
& && \rho_S \in [0, 1], \quad S \in \mathcal{K}.
\end{aligned} \tag{7}$$

Also define  $IC_{FLP}$  to be equal to 1 for singleton sets. The minimum broadcast rate of an index coding problem on the side information graph  $\mathcal{G}$  is bounded from above by  $IC_{FLP}(\mathcal{G})$ .

The proof of this theorem is going to be evident at the end of section IV and section V.

1) *Examples of better performance for our scheme:* While it had been evident that our proposed bounds are at least as good as the previous bounds, we now give explicit example of graphs where our upper bounds are strictly smaller than all previous upper bounds. Consider the index coding problem described by the graph in fig. 2. For this graph, the index coding based on the fractional local chromatic number (LP relaxation of theorem 1) has broadcast rate 4, the index coding based on the fractional partial clique clique covering (LP relaxation of eq. (3)) has broadcast rate 11/3 and our proposed scheme (theorem 4) has broadcast rate 7/2.

The proposed scheme of theorem 5 is also better than the corresponding recursive scheme proposed in [2, theorem 4]. For the graph in fig. 3 our scheme is strictly better than their scheme. The index coding broadcast rates for the graph are 3 and 7/2 for our proposed scheme and the scheme in [2, theorem 4], respectively.

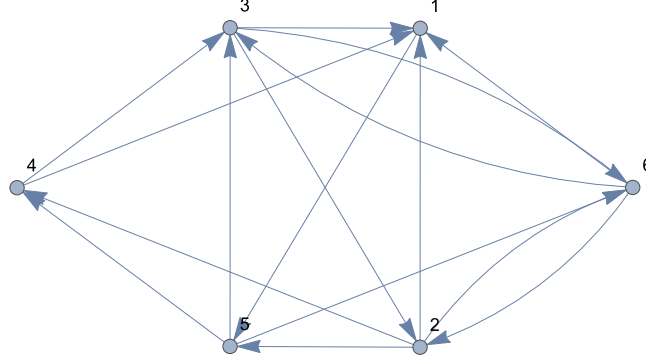


Fig. 3: Example of an Index Coding problem where the scheme in [2, theorem 4] are strictly larger than the proposed scheme for the fractional case.

*Remark (Codes with small alphabet size):* Consider the construction of linear index codes using local chromatic number of the complement graph in [14]. Assume that the local chromatic number  $\chi_l$  corresponds to a coloring with  $m$  colors. Then, we require an  $[m, \chi_l + 1]$ -MDS code. Using a Reed-Solomon code we can construct a  $[m, \chi_l + 1]$ -MDS code on a field of size  $m$ . Here, we note that, instead of using the generator matrix of a  $[m, \chi_l + 1]$ -MDS code the parity check matrix of any linear code of size  $m$  and minimum distance  $\chi_l + 2$  would work. Thus, when restricted to using a small alphabet size (say  $q$ ), we have the following upper-bound on the size of the code using the Gilbert-Varshamov bound,  $m - \log_q A_q(m, \chi_l + 2) = m - \log_q \left( \frac{q^m}{\sum_{j=0}^{\chi_l+1} \binom{m}{j} (q-1)^j} \right) = \log_q \left( \sum_{j=0}^{\chi_l+1} \binom{m}{j} (q-1)^j \right)$ .

### III. FURTHER RESULTS: EXTENSION TO $k$ -GIC GRAPHS

In the previous section, we modified the scheme in [14] by substituting a 0-partial clique cover of  $\mathcal{G}$  with a  $k$ -partial clique cover for any  $k > 0$ . We can further extend this procedure to include the  $n$ -GIC graphs defined in [15]. To that end, we extend the definition of  $n$ -GIC graphs to  $(k, n)$ -GIC, described below ( $n$ -GIC of [15] is  $(0, n)$ -GIC of our definition).

Consider a directed graph with  $N$  vertices having the following properties:

1. A set of  $n$  vertices, denoted by  $V_I$ , such that for any vertex  $v_i$  and at least  $n - k - 1$  vertices  $v_j \in V_I$  there is a path from  $v_i$  to  $v_j$  which does not include any other vertex of  $V_I$ . We call  $V_I$  the *inner vertex set*, and let  $V_I = \{v_1, v_2, \dots, v_n\}$ . The vertices of  $V_I$  are referred to as *inner vertices*.
2. Due to the above property, we can always find a *directed rooted tree*, (denoted by  $T_i$ ) with maximum number of leaves in  $V_I$  and root vertex  $v_i$ , having at least  $n - k - 1$  other vertices in  $V_I \setminus \{v_i\}$  as leaves. The trees may be non-unique.

Denote the union of all selected  $n$  trees as  $D_n \triangleq \bigcup_{v_i \in V_I} T_i$ . If  $D_n$  satisfies two conditions (to be defined shortly), we call it a  $(k, n)$ -GIC structure (denoted as a  $(k, n)$ -GIC sub-digraph:  $D_n = (V(D_n), E(D_n))$ , where  $|V(D_n)| = N$ ). Now we define a type of cycle and a type of path.

**Definition 5** (I-cycle). *A cycle that includes only one inner vertex  $i \in V_I$  is an I-cycle.*



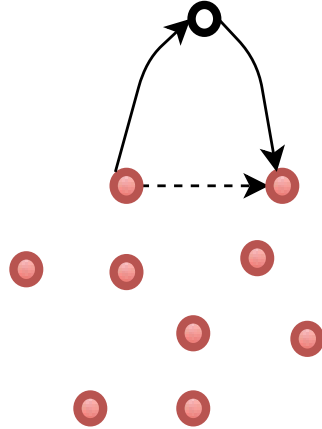


Fig. 4: Example of  $k$ -GIC [15]. The vertices shown in red are the inner vertices. The inner vertex set contains all possible edges (not shown) within the subgraph except the dotted edge.

**Definition 6** (P-path). *A path in which only the first and the last vertices are from  $V_I$ , and they are distinct, is a P-path.*

The conditions for  $D_n$  to be qualified as a  $(k, n)$ -GIC are as follows:

- 1) *Condition 1:* There is no  $I$ -cycle.
- 2) *Condition 2:* For all ordered pairs of inner vertices  $(v_i, v_j)$ ,  $i \neq j$ , there is only one  $P$ -path from  $i$  to  $j$ .

Figure 4 shows the example of a  $k$ -GIC as defined in [15]. It is an “almost” complete graph on the inner vertex set i.e. if all the  $P$ -paths are replaced with edges then the inner vertices form a 0-partial clique. Our definition of  $(k, n)$ -GIC extends the definition in [15] such that when all the  $P$ -paths are replaced with edges, the inner vertex set  $V_I$  forms a  $k$ -partial clique.

#### A. Code Construction

The following theorem allows us to construct an index coding scheme.

**Theorem 6.** *If a vertex  $v \in V \setminus V_I$  belongs to trees  $T_i$  and  $T_j$ ,  $i \neq j$ , then all the non-inner nodes on the subtree of  $T_i$  rooted at  $v$  also belong to  $T_j$ .*

Note that, although theorem 6 is similar to [15, lemma 3] it is different in that it applies to  $(n, k)$ -GIC in contrast to [15] where applies only to  $(n, 0)$ -GIC. The proof of this theorem is deferred to Appendix B. In particular in Appendix B, we show that the proof in [15, Lemma 3] still applies for the  $(k, n)$ -GIC.

Denote the non-inner vertices of the subtree rooted at vertex  $v \in V \setminus V_I$  as  $D_n(v)$ . Let  $[v_1 \ v_2 \ \dots \ v_n]$  be the generator matrix of a  $[n, k+1]$ -MDS code. Then the index code is defined below:

1.  $\mathbf{w}_I = \sum_{v_j \in V_I} \mathbf{v}_j x_j$ .
2. Let  $\mathbf{x}_{N(v_j, D_n)} \in \mathbb{F}_q^{|N(v_j, D_n)|}$  denote the vector consisting of the data corresponding to the nodes in  $N(v_j, D_n)$ . A vector  $\mathbf{w}_j \in \mathbb{F}_q^{\min\{|N(v_j, D_n)|, k+1\}}$  corresponding to all vertices  $v_j \in V \setminus V_I$  is transmitted,

where,

$$\mathbf{w}_j = \begin{cases} \sum_{v_l \in \{N(v_j, D_n) \cap V_I\}} \mathbf{v}_l(x_j + x_l) + \sum_{v_l \in \{N(v_j, D_n) \setminus V_I\}} \mathbf{u}_l(x_j + x_l) & \text{if } |N(v_j, D_n)| \geq k+1 \\ \mathbf{1}x_j + \mathbf{x}_{N(v_j, D_n)} & \text{if } |N(v_j, D_n)| < k+1 \end{cases} \quad (8)$$

where  $\mathbf{u}_l \in \mathbb{F}_q^{\min\{|N(v_j, D_n)|, k+1\}}$  are chosen as described in algorithm 1. Algorithm 1 first assigns a vector  $\mathbf{u}_l$  to all non-inner nodes  $v_l \in D_n$  starting from the nodes at the bottom of a tree  $T_i \in D_n$ . The vector  $\mathbf{u}_l$  corresponding to the vertex  $v_l$  depends on its children  $\{\mathbf{u}_l : v_l \in D_n(v_l)\}$ . To show that the algorithm works, we need to find a vertex for which all out-neighbors are outside  $S$ . It is easy to see from theorem 6 that this always holds.

**Data:** trees  $T_1, T_2, \dots, T_n$

**Result:**  $\mathbf{u}_j \in \mathbb{F}_q^{\min\{|N(v_j, D_n)|, k+1\}}$  for  $v_j \in V \setminus V_I$  and  $\mathbf{u}_j \in \mathbb{F}^{k+1}$  for  $v_j \in V_I$

```

1  $\mathbf{u}_i = \mathbf{v}_i$  for all  $v_i \in V_I$ 
2  $S = V \setminus V_I$ 
3 while  $|S| > 0$  do
4   Find a vertex in  $v_i \in S : N(v_i, D_n) \subseteq \overline{S}$ 
5    $\mathbf{u}_i = - \sum_{j: v_j \in N(v_i, D_n)} \mathbf{u}_j$ 
6    $S = S \setminus \{v_i\}$ 
7 end
```

**Algorithm 1:** Selecting the vectors  $\mathbf{u}_j, j \in V \setminus V_I$

### B. Decoding

It is easy to see that all the non-inner vertices  $j \in V \setminus V_I$  can recover their data  $x_j$ . We show that  $v_i \in V_I$  can also recover  $x_i$ . Define  $\mathbf{w}_j'$  corresponding to the transmitted vector  $\mathbf{w}_j$  for  $v_j \in V \setminus V_I$  as follows,

$$\mathbf{w}_j' = \begin{cases} \mathbf{w}_j & \text{if } |N(v_j, D_n)| \geq k+1 \\ [\mathbf{u}_{c_1} \ \mathbf{u}_{c_2} \ \dots \ \mathbf{u}_{c_r}] \mathbf{w}_j & \text{if } |N(v_j, D_n)| < k+1 \end{cases} \quad (9)$$

where  $\{v_{c_1}, v_{c_2}, \dots, v_{c_r}\} = N(v_j, D_n)$ . Denote by  $T(v)$  the subtree rooted at vertex  $v$  in tree  $T$ . For the non-inner child  $v$  of vertex  $v_i$  compute the following,

$$\mathbf{w}(v) = \sum_{v_l \in T_i(v) \setminus V_I} \mathbf{w}_l' \quad (10)$$

Note that,  $\mathbf{w}(v)$  only contains terms of the form  $\mathbf{u}_l x_l$  for  $\{l : v_l \in V_I \text{ or } v_l \in D_n(v_i)\}$ . Now, consider  $\mathbf{w}_I - \sum_{v \in D_n(v_i)} \mathbf{w}(v)$ . The only terms left in this are  $\mathbf{u}_l x_l$  for  $\{l : v_l \in V_I \setminus T_i\}$ . Since there are at most  $k$  such terms and  $[\mathbf{v}_1 \ \mathbf{v}_2 \ \dots \ \mathbf{v}_n]$  is the generator matrix of a  $[n, k+1]$ -MDS code, vertex  $v_i$  can decode  $x_i$ .

#### IV. ACHIEVABILITY SCHEME (PROOF OF THEOREM 3)

We now describe an index coding scheme to achieve a broadcast rate equal to the optimal solution of the program in eq. (5).

Assume without loss of generality that  $\mathcal{S}_1 = [n_1]$ ,  $\mathcal{S}_2 = [n_1 + 1, n_1 + n_2]$ ,  $\dots$ ,  $\mathcal{S}_t = [\sum_{j \in [t-1]} n_j + 1, \sum_{j \in [t]} n_j]$  be the partial cliques selected in eq. (5). Let  $k_j := k_{\mathcal{S}_j}$ . Assume that the optimum value of eq. (5) is  $m$ . Then  $\max_j (k_j + 1) \leq m \leq \sum_j (k_j + 1)$ . We construct  $[n_j, k_j + 1]$ -MDS codes for each  $j \in [t]$ . Let  $G_j$  denote the generator matrices for these codes. Thus,

$$G_j = \begin{bmatrix} 1 & 1 & \dots & 1 \\ \alpha_{j,1} & \alpha_{j,2} & \dots & \alpha_{j,n_j} \\ & & \ddots & \\ \alpha_{j,1}^{k_j} & \alpha_{j,2}^{k_j} & \dots & \alpha_{j,n_j}^{k_j} \end{bmatrix} \quad (11)$$

where  $\alpha_{j,1}, \alpha_{j,2}, \dots, \alpha_{j,n_j} \in \mathbb{F}_q$  for all  $j \in [t]$ .

Let  $k^j = \sum_{l=1}^j (k_l + 1)$ . Assume that  $\Phi$  is the generator matrix of an  $[k^t, m]$ -MDS code. Let  $\Phi_j = \Phi_{[k^{j-1}+1, k^j]}$ . Denote,  $x_j^i = \min\{|\mathcal{S}_j \cap \overline{N(v_i, \mathcal{G})}|, (k_{\mathcal{S}_j} + 1)\}$ ,  $i \in [n]$ ,  $j \in [t]$ .

Finally we construct the following matrix  $G$ ,

$$G = [\mathbf{u}_1 \ \mathbf{u}_2 \ \dots \ \mathbf{u}_n] = [\Phi_1 G_1 \ \Phi_2 G_2 \ \dots \ \Phi_t G_t]_{m \times n} \quad (12)$$

where  $n = \sum_j n_j$  is the number of vertices in  $\mathcal{G}$ .

Now assigning the vector  $\mathbf{u}_i$  to vertex  $i$ , we show that this assignment satisfies the interference alignment condition in eq. (1). Note that this scheme is a natural generalization of the localized coloring scheme where the matrix  $G_j$  is the generator matrix of a repetition code and  $k^t = m$ .

**Lemma 7.** *Let  $H_j$  be a  $(k_j + 1) \times (k_j + 1)$  submatrix of  $G_j$ . Then for a large enough field  $\mathbb{F}_q$  there exist  $\alpha_{j,i} \in \mathbb{F}_q$ ,  $j \in [t]$ ,  $i \in [n_j]$  such that  $\hat{G} = [\Phi_1 H_1 \ \Phi_2 H_2 \ \dots \ \Phi_t H_t]$  is an MDS matrix for any set of sub-matrices  $H_j$ .*

*Proof:* Let  $\Phi = [\mathbf{v}_1 \ \mathbf{v}_2 \ \dots \ \mathbf{v}_{k^t}]$  and let  $U(s)$  denote any  $m \times s$  sub-matrix of  $\Phi_{\geq 2} := [\Phi_2 \ \dots \ \Phi_t]$ . Since  $\Phi$  is MDS,  $[\mathbf{v}_{i_1} \ \mathbf{v}_{i_2} \ \dots \ \mathbf{v}_{i_r} \ U(m-r)]$  must be full-rank for all  $\{i_1, \dots, i_r\} \subseteq [k_1 + 1]$ . Consider any vector  $\mathbf{w} \in \mathbb{F}_q^m$ ,

$$\mathbf{w} = [\mathbf{v}_{i_1} \ \mathbf{v}_{i_2} \ \dots \ \mathbf{v}_{i_r} \ U(m-r)] \begin{bmatrix} \mathbf{a}_r \\ \mathbf{a}_{r+1} \end{bmatrix} \quad (13)$$

where  $\mathbf{a}_r \in \mathbb{F}_q^r$  and  $\mathbf{a}_{r+1} \in \mathbb{F}_q^{m-r}$ . We show that there exist  $\alpha_{1,i} \in \mathbb{F}_q$ ,  $i \in [n_1]$  such that  $\mathbf{w}$  can also be represented as a linear combination of column vectors in  $G' := [\Phi_1 H_1 \ U(m-r)]$  where,

$$H_1 = \begin{bmatrix} 1 & 1 & \dots & 1 \\ \alpha_{1,i_1} & \alpha_{1,i_2} & \dots & \alpha_{1,i_r} \\ & & \ddots & \\ \alpha_{1,i_1}^{k_1} & \alpha_{1,i_2}^{k_1} & \dots & \alpha_{1,i_r}^{k_1} \end{bmatrix}$$

for any  $\{i_1, i_2, \dots, i_r\} \subseteq [n_1]$ , i.e.

$$\mathbf{w} = G' \begin{bmatrix} \mathbf{d}_r \\ \mathbf{d}_{r+1} \end{bmatrix} \quad (14)$$

for some  $\mathbf{d}_r \in \mathbb{F}_q^r$  and  $\mathbf{d}_{r+1} \in \mathbb{F}_q^{m-r}$ .

Since,  $[\mathbf{v}_{i_1} \mathbf{v}_{i_2} \dots \mathbf{v}_{i_r} U(m-r)]$  is full-rank, we have,

$$[\mathbf{v}'_{i'_1} \mathbf{v}'_{i'_2} \dots \mathbf{v}'_{i'_{k_1+1-r}}] = [\mathbf{v}_{i_1} \mathbf{v}_{i_2} \mathbf{v}_{i_r} U(m-r)] \begin{bmatrix} B_r \\ B_{r+1} \end{bmatrix}, \quad (15)$$

where  $B_r \in \mathbb{F}_q^{r \times (k_1+1-r)}$  and  $B_{r+1} \in \mathbb{F}_q^{(m-r) \times (k_1+1-r)}$  and  $\{i'_1, i'_2, \dots, i'_{k_1+1-r}\} = [k_1+1] \setminus \{i_1, i_2, \dots, i_r\}$ . Thus, combining eqs. (13) to (15), we have,

$$\begin{aligned} & [\mathbf{v}_{i_1} \mathbf{v}_{i_2} \dots \mathbf{v}_{i_r} U(m-r)] \cdot \\ & \left( \begin{bmatrix} [I_r \ B_r] P_\sigma H_1 \mathbf{d}_r \\ B_{r+1} H'_1 \mathbf{d}_r + \mathbf{d}_{r+1} \end{bmatrix} - \begin{bmatrix} \mathbf{a}_r \\ \mathbf{a}_{r+1} \end{bmatrix} \right) = \mathbf{0} \end{aligned} \quad (16)$$

where,

$$H'_1 = \begin{bmatrix} \alpha_{1,i_1}^{r+1} & \alpha_{1,i_2}^{r+1} & \dots & \alpha_{1,i_r}^{r+1} \\ \alpha_{1,i_1}^{r+2} & \alpha_{1,i_2}^{r+2} & \dots & \alpha_{1,i_r}^{r+2} \\ \vdots & \vdots & \ddots & \vdots \\ \alpha_{1,i_1}^{k_1} & \alpha_{1,i_2}^{k_1} & \dots & \alpha_{1,i_r}^{k_1} \end{bmatrix}$$

and  $P_\sigma$  is the permutation matrix for  $\sigma = (1 \rightarrow i_1, 2 \rightarrow i_2, \dots, r \rightarrow i_r, (r+1) \rightarrow i'_1, r+2 \rightarrow i'_2, \dots, (k+1) \rightarrow i'_{k+1-r})$ . For the solution in eq. (16) to exist for all  $\mathbf{a}_r \in \mathbb{F}_q^r, \mathbf{a}_{r+1} \in \mathbb{F}_q^{k-r+1}$  we must have,  $\det([I_r \ B_r] P_\sigma H_1) \neq 0$  i.e.,

$$\det \begin{bmatrix} \alpha_{1,i_1}^{i_1-1} + \mathbf{g}(\alpha_{1,i_1}) \mathbf{b}_1 & \alpha_{1,i_2}^{i_1-1} + \mathbf{g}(\alpha_{1,i_2}) \mathbf{b}_1 & \dots & \alpha_{1,i_r}^{i_1-1} + \mathbf{g}(\alpha_{1,i_r}) \mathbf{b}_1 \\ \alpha_{1,i_1}^{i_2-1} + \mathbf{g}(\alpha_{1,i_1}) \mathbf{b}_2 & \alpha_{1,i_2}^{i_2-1} + \mathbf{g}(\alpha_{1,i_2}) \mathbf{b}_2 & \dots & \alpha_{1,i_r}^{i_2-1} + \mathbf{g}(\alpha_{1,i_r}) \mathbf{b}_2 \\ \vdots & \vdots & \ddots & \vdots \\ \alpha_{1,i_1}^{i_r-1} + \mathbf{g}(\alpha_{1,i_1}) \mathbf{b}_r & \alpha_{1,i_2}^{i_r-1} + \mathbf{g}(\alpha_{1,i_2}) \mathbf{b}_r & \dots & \alpha_{1,i_r}^{i_r-1} + \mathbf{g}(\alpha_{1,i_r}) \mathbf{b}_r \end{bmatrix} \neq 0 \quad (17)$$

where  $\mathbf{g}(\alpha) = [\alpha^{i'_1-1} \alpha^{i'_2-1} \dots \alpha^{i'_{k_1+1-r}-1}]$  and  $B_r = [\mathbf{b}_1 \ \mathbf{b}_2 \ \dots \ \mathbf{b}_r]^T$ .

The polynomial in the RHS of eq. (17) has degree at most  $k_1$  for all the variables  $\alpha_{1,i_1}, \alpha_{1,i_2}, \dots, \alpha_{1,i_r}$ . Thus, by increasing the size of the field  $\mathbb{F}_q$  we can make sure that there exist  $\alpha_{j,i}$  for all  $j \in [t]$  and  $i \in [n_j]$  such that eq. (17) holds.

Now, repeating the above argument  $t$  times we can say that  $[\Phi_1 H_1 \Phi_2 H_2 \ \Phi_3 H_3 \ \dots \ \Phi_t H_t]$  is MDS for all sets of  $H_j \in \mathbb{F}_{q^r}^{(k_j+1) \times (k_j+1)}$  submatrices of  $G_j$ .

Whereas a loose upper-bound (sufficient) on the field size becomes,

$$q \leq \max_{j \in [t]} \sum_{r=1}^{k_j+1} k_j \binom{n-n_j}{m-r} \binom{n_j}{r-1} + n_j \quad (18)$$

Now our main theorem is evident. ■

**Theorem 8.** For any vertex  $v_i \in V(\mathcal{G})$  we have,

$$\mathbf{u}_i \notin \text{span}(\mathbf{u}_{\overline{N(v_i, \mathcal{G}) \setminus v_i}}) \quad (19)$$

*Proof:* Consider a vertex  $v_i$  in graph  $\mathcal{G}$ , with non-neighbors  $P_1, P_2, \dots, P_t$  such that  $P_j \subseteq \{i : v_i \in \mathcal{S}_j\}$ . Assume, wlog that  $v_i \in \mathcal{S}_1$ . Then,  $|P_1| \leq k_1$ . Note that, for any  $P_j : |P_j| \geq k_j + 1$  there exists a set of  $P'_j \subseteq P_j : |P'_j| = k_j + 1$  and  $\text{span}(\mathbf{u}_{P_j}) = \text{span}(\mathbf{u}_{P'_j})$ . Let  $P'_j = P_j$  for  $P_j : |P_j| \leq k_j + 1, j \neq 1$

and  $P'_1 = P_1 \cup i$ . Therefore, we have  $\text{span}(\{\mathbf{u}_{P_j}\}_j) = \text{span}(\{\mathbf{u}_{P'_j}\}_j)$ . Now, since  $\{\mathbf{u}_{P'_j}\}_j$  is full-rank (lemma 7) its easy to see that eq. (19) follows. ■

Note that, in the proof of theorem 8 we do not need the matrix  $\hat{G}$  in lemma 7 to be MDS. Instead we only need a class of  $n$  different subsets of column vectors of  $G$ , each of size at most  $m$ , to be linearly independent. Thus the upper bound on the size of the alphabet in eq. (18) is very loose and we can show that an alphabet of size  $O(n)$  should suffice (details omitted).

The optimal index coding solution in eq. (5) uses at most 2 levels of MDS codes  $\Phi$  and  $G_j, j \in [t]$ . If we recursively use the above method on the  $k$ -partial cliques corresponding to  $G_j$  then we can further reduce the index coding broadcast rate in eq. (5). The achievement scheme for the recursive linear program can be easily obtained by replacing  $G_j$  in eq. (12) with the matrix obtained for  $IC_{LP}(\mathcal{G}|_S)$ . The matrix for  $IC_{LP}(\mathcal{G}|_S)$  can be obtained recursively in a similar manner. The integer program corresponding to the recursive scheme is,

$$\begin{aligned} & \text{minimize} && t \\ & \text{subject to} && \sum_{S \in \mathcal{K}} \min\{|\mathcal{S} \cap \overline{N(v, \mathcal{G})}|, IC_{LP}(\mathcal{G}|_S)\} \rho_S \leq t, \\ & && v \in V(\mathcal{G}) \\ & \text{and} && \sum_{S \in \mathcal{K}: v \in S} \rho_S \geq 1, \quad v \in V(\mathcal{G}) \\ & && \rho_S \in \{0, 1\}, \quad S \in \mathcal{K} \end{aligned} \tag{20}$$

where  $IC_{LP}(\mathcal{G})$  denotes the optimum of the above integer program for the graph  $\mathcal{G}$  and  $\mathcal{G}|_S$  denotes the graph  $\mathcal{G}$  restricted to the subset  $S$ . Note that  $IC_{LP}(\mathcal{G}|_{\{i\}}) := 1, i \in [n]$  and the proof in lemma 7 and theorem 8 apply to the scheme for eq. (20) as well.

## V. FRACTIONAL LOCAL PARTIAL CLIQUE COVER (PROOF OF THEOREM 4)

In this section we show how to design an encoding scheme to achieve the rate given by the linear programming optimal solution of eq. (6).

Since all the coefficients of the LP in eq. (6) are integers, the optimal solution  $\rho_S^*$  must be rational. Assume that in optimal solution to eq. (6) all partial cliques  $S$  for which  $\rho_S > 0$  are  $S_1, S_2, \dots, S_t$  and  $|\mathcal{S}_j| = n_j$ . Let  $k_j := k_{S_j}$ . Also assume that all  $\rho_{S_j}, j \in [t]$  have a common denominator  $\Delta$  such that  $\rho_{S_j} = N_j/\Delta$ . Assume that the optimum value to the optimization of eq. (6) is  $m^*$ . Construct a  $[n_j, k_j + 1]$ -MDS code for each  $j$ . Let  $G_j, j \in [t]$  be the generator matrix for these codes. Thus,

$$G_j = \begin{bmatrix} 1 & 1 & \dots & 1 \\ \alpha_{j,1} & \alpha_{j,2} & \dots & \alpha_{j,n_j} \\ & & \ddots & \\ \alpha_{j,1}^{k_j} & \alpha_{j,2}^{k_j} & \dots & \alpha_{j,n_j}^{k_j} \end{bmatrix} \tag{21}$$

where  $\alpha_{j,1}, \alpha_{j,2}, \dots, \alpha_{j,n_j} \in \mathbb{F}$  for all  $j \in [t]$ . Let  $G'_j := \sum_{l=1}^{N_j} Q_{ll} \otimes G_j$ , where  $Q_{rs}$  is a  $N_j \times N_j$  matrix with the only non-zero entry being  $Q_{rs}(r, s) = 1$  and  $\otimes$  denotes the matrix tensor product. Let  $k^j = \sum_{l=1}^j N_j(k_l + 1)$  and let  $\Phi$  be the generator matrix of an  $[k^t, m^* \Delta]$ -MDS code. Let  $\Phi_j = \Phi_{[k^{j-1}+1, k^j]}$ . Finally we construct a matrix  $G$ :

$$G = [\mathbf{u}_{1,1} \dots \mathbf{u}_{1,(N_1 n_1)} \dots \mathbf{u}_{t,1} \dots \mathbf{u}_{t,N_t n_t}] \tag{22}$$

$$= [\Phi_1 G'_1 \quad \Phi_2 G'_2 \quad \dots \quad \Phi_t G'_t]_{(\Delta m^*) \times (\Delta n)} \tag{23}$$

Note that  $\sum_{j=1}^t n_j N_j = \Delta n$ . Now each vertex has to be assigned  $\Delta$  column vectors from  $G$ . Assume wlog that a vertex belongs to partial cliques  $S_1, S_2, \dots, S_r$ ,  $r \leq t$ . Assume that the column vector in  $G_j$ ,  $j \in [r]$  corresponding to the vertex is  $\begin{bmatrix} 1 & \alpha_{j,c_j} & \dots & \alpha_{j,c_j}^{k_j} \end{bmatrix}^T$ ,  $c_j \in [n_j]$ . Then the vectors assigned to the vertex are  $\mathbf{u}_{j,c_j}, \mathbf{u}_{j,c_j+n_j}, \mathbf{u}_{j,c_j+2n_j}, \dots, \mathbf{u}_{j,c_j+(N_j-1)n_j}$  for  $j \in [r]$ .

Now using arguments similar to lemma 7, the interference alignment condition can be satisfied for this case too.

To construct an achievability for the fractional relaxation of the recursive scheme (eq. (24)), we can recursively use the scheme for eq. (6).

$$\begin{aligned}
& \text{minimize} && t \\
& \text{subject to} && \sum_{S \in \mathcal{K}} \min_{v \in V(\mathcal{G})} \{ |\mathcal{S} \cap \overline{N(v, \mathcal{G})}|, IC_{FLP}(\mathcal{G}|_S) \} \rho_S \leq t, \\
& \text{and} && \sum_{S \in \mathcal{K}: v \in S} \rho_S \geq 1, \quad v \in V(\mathcal{G}) \\
& && \rho_S \in [0, 1], \quad S \in \mathcal{K}
\end{aligned} \tag{24}$$

where  $IC_{FLP}(\mathcal{G})$  denotes the optimum value of the above linear program for graph  $\mathcal{G}$ . Let a vertex  $i \in [n]$  belong to  $r$  recursive index coding problems  $IC_{LP}(\mathcal{G}), IC_{LP}(\mathcal{G}|_{S_1}), \dots, IC_{LP}(\mathcal{G}|_{S_r})$ ,  $\mathcal{G} \supset S_1 \supset S_2 \dots \supset S_r$ . Thus, for the achievability scheme above, it is assigned  $r$  coefficients corresponding to the  $r$  problems,  $\rho_1 = N^1/\Delta^1, \dots, \rho_r = N^r/\Delta^r$  and assigned  $\Delta^1 \times \dots \times \Delta^r$  column vectors.

Note that, we need the size of the node,  $\ell$ , to be at least  $\Delta^1 \times \dots \times \Delta^r$ , for the scheme to work.

## REFERENCES

- [1] Noga Alon, Eyal Lubetzky, Uri Stav, Amit Weinstein, and Avinatan Hassidim. Broadcasting with side information. In *Foundations of Computer Science, 2008 (FOCS'08), 49th Annual Symposium on*, pages 823–832. IEEE, 2008.
- [2] Fatemeh Arbabjolfaei and Young-Han Kim. Local time sharing for index coding. In *Information Theory (ISIT), 2014 IEEE International Symposium on*, pages 286–290. IEEE, 2014.
- [3] Ziv Bar-Yossef, Yitzhak Birk, TS Jayram, and Tomer Kol. Index coding with side information. *Information Theory, IEEE Transactions on*, 57(3):1479–1494, 2011. Preliminary version in FOCS 2006.
- [4] Yitzhak Birk and Tomer Kol. Informed-source coding-on-demand (iscod) over broadcast channels. In *INFOCOM'98. Seventeenth Annual Joint Conference of the IEEE Computer and Communications Societies. Proceedings. IEEE*, volume 3, pages 1257–1264. IEEE, 1998.
- [5] Anna Blasiak, Robert Kleinberg, and Eyal Lubetzky. Lexicographic products and the power of non-linear network coding. In *Foundations of Computer Science (FOCS), 2011 IEEE 52nd Annual Symposium on*, pages 609–618. IEEE, 2011.
- [6] Michelle Effros, Salim El Rouayheb, and Michael Langberg. An equivalence between network coding and index coding. *Information Theory, IEEE Transactions on*, 61(5):2478–2487, 2015.
- [7] Salim El Rouayheb, Alex Sprintson, and Costas Georgiades. On the index coding problem and its relation to network coding and matroid theory. *Information Theory, IEEE Transactions on*, 56(7):3187–3195, 2010.
- [8] Paul Erdős, Zoltán Füredi, András Hajnal, Péter Komjáth, Vojtech Rödl, and Ákos Seress. Coloring graphs with locally few colors. *Discrete mathematics*, 59(1):21–34, 1986.
- [9] W. Haemers. An upper bound on the shannon capacity of a graph. *Algebraic methods in Graph Theory*, 25:267–272, 1978.
- [10] Young-Han Kim. All index coding problems up to  $n=5$  messages. <http://circuit.ucsd.edu/~yhk/indexcoding.html>. Accessed: 2016-02-17.
- [11] Michael Langberg and Alex Sprintson. On the hardness of approximating the network coding capacity. In *Information Theory, 2008. ISIT 2008. IEEE International Symposium on*, pages 315–319. IEEE, 2008.
- [12] Eyal Lubetzky and Uri Stav. Nonlinear index coding outperforming the linear optimum. *Information Theory, IEEE Transactions on*, 55(8):3544–3551, 2009.
- [13] Karthikeyan Shanmugam and Alexandros G. Dimakis. Connections between index coding, locally repairable codes and the multiple unicast problem. *personal communication*, 2014.

- [14] Karthikeyan Shanmugam, Alexandros G Dimakis, and Michael Langberg. Local graph coloring and index coding. In *Information Theory Proceedings (ISIT), 2013 IEEE International Symposium on*, pages 1152–1156. IEEE, 2013.
- [15] Chandra Thapa, Lawrence Ong, and Sarah J Johnson. Generalized interlinked cycle cover for index coding. *arXiv preprint arXiv:1504.04806*, 2015.

## APPENDIX

### A. A Simpler Achievability Scheme for the Recursive Scheme in [2, theorem 4]

Consider the following integer program proposed in [2],

$$\begin{aligned}
 & \text{minimize} && t \\
 & \text{subject to} && \sum_{S \in \mathcal{K}} (k_S + 1) \rho_S \leq t, \\
 & && v \in V(\mathcal{G}) \\
 & \text{and} && \sum_{S \in \mathcal{K}: v \in S} \rho_S \geq 1, \quad v \in V(\mathcal{G}) \\
 & && \rho_S \in [0, 1], \quad S \in \mathcal{K}.
 \end{aligned} \tag{25}$$

We propose a simple scheme corresponding to the integer program in eq. (25). Assume that the optima corresponding to the integer program is  $m$ , the partial cliques selected are  $\mathcal{S}_1, \mathcal{S}_2, \dots, \mathcal{S}_t$ . Let  $n_j := |\mathcal{S}_j|$  and  $k_j := k_{\mathcal{S}_j}$ . Let the cliques selected corresponding to the vertex  $v_i$  have indices  $\{c_1^i, c_2^i, \dots, c_{t_i}^i\} \subseteq [t]$ .

We construct the following matrix  $G$ ,

$$G = [\mathbf{u}_1 \ \mathbf{u}_2 \ \dots \ \mathbf{u}_n] = [\Phi_1 G_1 \ \Phi_2 G_2 \ \dots \ \Phi_t G_t]_{m \times n} \tag{26}$$

where  $n = \sum_j n_j$  is the number of vertices in  $\mathcal{G}$ ,  $G_j, j \in [t]$  are  $[n_j, k_j + 1]$ -MDS matrices, and  $\Phi_j$  are  $m \times (k_j + 1)$  matrices defined as follows.

Let  $\Phi$  denote a set of  $m$  linearly independent vectors of dimension  $m$ . We construct the matrix  $\Phi_j$  such that its column vectors are vectors in  $\Phi$  and the column vectors for the matrices  $\Phi_{c_1^i}, \dots, \Phi_{c_{t_i}^i}$  are full rank for all  $i$ . Note that such an assignment is possible because  $m = \max_i \sum_{j=1}^{t_i} (k_{c_j^i} + 1)$ .

Note that, in this case, the upper bound on the alphabet size is  $\max_{j \in [t]} n_j$  corresponding to the largest alphabet size needed for constructing  $G_j$ .

The achievability scheme for the recursive scheme in [2, theorem 4] can be constructed from eq. (25) analogous to the way for which we construct the scheme for eq. (20) from eq. (5). And the extension to linear programs is also done in an analogous manner.

### B. Proof of theorem 6

Denote the leaves of subtree of the tree  $T$  rooted at vertex  $v \in V(T)$  as  $L(v, T)$  and the leaves of the tree  $T$  as  $L(T)$ .

**Lemma 9.** *If a vertex  $v \in V(T_i)$  and  $v \in V(T_j)$ ,  $i \neq j$  and  $v \in V_1$  then  $L(v, T_j), L(v, T_i) \subseteq V_1 \setminus \{v_i, v_j\}$*

*Proof:* If the vertex  $v_j \in L(v, T_i)$ , then there exists a path from vertex  $v$  to  $v_j$  in the tree  $T_i$ . However, in the tree  $T_j$ , there is a path from vertex  $v_j$  to  $v$ . Thus in the sub-digraph  $D_n$ , we obtain a path from vertex  $v$  to  $v_j$  (via  $T_i$ ) and vice versa (via  $T_j$ ). As a result, an  $I$ -cycle containing  $v_j$  is present. This contradicts the condition 1 (i.e., no  $I$ -cycle) for a  $D_n$ . Hence,  $v_j \notin L(v, T_i)$ . In other words,  $L(v, T_i) \subseteq V_1 \setminus \{v_i, v_j\}$ . Similarly,  $L(v, T_j) \subseteq V_1 \setminus \{v_i, v_j\}$ . ■

**Lemma 10.** *If a vertex  $v \in V(T_i)$  and  $v \in V(T_j)$ ,  $i \neq j$  and  $v \in V_1$  then  $L(v, T_i) = L(v, T_j)$*

*Proof:* From lemma 9,  $L(v, T_i)$  is a subset of  $V_I \setminus \{v_i, v_j\}$ . Now pick a vertex  $v_c$  belongs to  $V_I \setminus \{v_i, v_j\}$  such that  $v_c \in L(v, T_i)$  but  $v_c \notin L(v, T_j)$  (such  $v_c$  exists since we suppose that  $L(v, T_i) \neq L(v, T_j)$ ). In tree  $T_i$ , there exists a directed path from the vertex  $v_i$  which includes the vertex  $v$ , and ends at the leaf vertex  $v_c$ . Let this path be  $P_{v_i \rightarrow v_c}(T_i)$ .

Now, suppose that in tree  $T_j$ , there exists a directed path from the vertex  $v_j$ , which doesn't include the vertex  $v$  (since  $v_c \notin L(v, T_j)$ ), and ends at the leaf vertex  $v_c$ . Let this path be  $P_{v_j \rightarrow v_c}(T_j)$ . However, in the digraph  $D_n$ , we can also obtain a directed path from the vertex  $v_j$  which passes through the vertex  $v$  (via  $T_j$ ), and ends at the leaf vertex  $v_c$  (via  $T_i$ ). Let this path be  $P_{v_j \rightarrow v_c}(D_n)$ . The paths  $P_{v_j \rightarrow v_c}(T_j)$  and  $P_{v_j \rightarrow v_c}(D_n)$  are different which indicates the existence of multiple  $P$ -paths from the vertex  $v_j$  to  $v_c$  in  $D_n$ , this contradicts the condition 2 for a  $D_n$ . Consequently,  $L(v, T_i) = L(v, T_j)$ .

Therefore, the only case left is  $v_c \notin L(T_j)$ . But since, the tree  $T_j$  must be such that it has the maximum number of leaves in  $V_I$  and there exists a tree that has more leaves than  $T_j$  this leads to a contradiction. ■

**Lemma 11.** *If a vertex  $v \in V(T_i)$  and  $v \in V(T_j)$ ,  $i \neq j$  and  $v \in V_I$  then the out-neighborhood of the vertex  $v$  must be same in both the trees i.e.  $N(v, T_i) = N(v, T_j)$*

*Proof:* Now we pick a vertex  $v_b$  such that, without loss of generality,  $v_b \in N(v, T_i)$  but  $v_b \notin N(v, T_j)$  (such  $v_b$  exists since we assumed that  $N(v, T_i) \neq N(v, T_j)$ ). Furthermore, we have two cases for  $v_b$ , which are (case 1)  $v_b \in L(v, T_i)$ , and (case 2)  $v_b \notin L(v, T_i)$ . Case 1 is addressed in lemma 10. On the other hand, for case 2, we pick a leaf vertex  $v_d \in L(v_b, T_i)$  such that there exists a path that starts from  $v$  followed by  $v_b$ , and ends at  $v_d$ , i.e.,  $\langle v, v_b, \dots, v_d \rangle$  exists in  $T_i$ . A path  $\langle v_j, \dots, v \rangle$  exists in  $T_j$ . Thus a path  $\langle v_j, \dots, v, v_b, \dots, v_d \rangle$  exists in  $D_n$ . From the first part of the proof, we have  $L(v, T_i) = L(v, T_j)$ , so  $v_d \in L(v, T_j)$ . Now in  $T_j$ , there exists a path from  $v_j$  to  $v_d$ , which includes vertex  $v$  followed by a vertex  $v_e$  such that  $v_e \in N(v, T_j)$  and  $v_e \neq v_b$  (as  $v_b \notin N(v, T_j)$ ), and the path ends at  $v_d$ , i.e.,  $\langle v_j, \dots, v, v_e, \dots, v_d \rangle$  which is different from  $\langle v_j, \dots, v, v_b, \dots, v_d \rangle$ . So multiple  $P$ -paths are observed at  $v_d$  from  $v_j$ . This contradicts condition 2 for a  $D_n$ . Consequently,  $N(v, T_i) = N(v, T_j)$ . ■